

# A WEB-Based Version of MedLEE: A Medical Language Extraction and Encoding System

Carol Friedman Ph.D.<sup>1,2</sup>, Lyudmila Shagina<sup>2</sup>, Socrates A. Socratous<sup>2</sup>, Xiao Zeng<sup>2</sup>

1 Department of Computer Science, Queens College of CUNY and  
Department of Medical Informatics, Columbia University, New York  
2 Department of Medical Informatics, Columbia University, New York

Clinical data comprises the heart of a Clinical Information System (CIS). More importantly, reliable access to the clinical data empowers a CIS so that automated processes can be performed on the data and thereby provide clinicians and health care administrators with invaluable information that could be used for decision support, research studies, quality assurance, outcomes assessment, and resource management. However, reliable access is often difficult to achieve because it requires structured data, and currently, the vast amount of clinical data that is available is typically in textual form as contained in patient reports.

One way to obtain structured data is to use natural language processing methods to automatically encode data that is in textual form. MedLEE [1], which has been developed for that purpose, is currently being used in daily operation at Columbia Presbyterian Medical Center (CPMC) to structure and encode clinical reports. MedLEE was initially developed to process radiological reports of the chest, and was subsequently expanded to handle mammography reports. These types of reports contain limited information and therefore are associated with relatively modest and well-defined vocabularies. The applications which initially accessed the output produced by MedLEE were decision support applications which typically require high precision.

MedLEE is currently being extended to the domain of discharge summaries in order to obtain another source of structured data and in order to support applications other than decision support. Expansion necessitated that more robust processing methods be developed because the domain includes a very comprehensive vocabulary, the representational model is broader, and the language model is much more complex. In addition, for some applications, high sensitivity is more important than high precision. Expansion into

other applications also required that a flexible interface be developed to accommodate the different needs of users.

In order to provide easy access to MedLEE, a Client-Server version running on a Unix-based platform has been implemented and a Common Gateway Interface (CGI) has been developed so that MedLEE can be called using a WEB browser. Data is entered via a WEB form which allows the user to specify text and input parameters that supply contextual information as well. After the form is submitted, the text is processed and structured output is returned to the user in Hypertext Markup Language (HTML), which is displayed by the WEB browser. Alternatively, authorized users can access MedLEE directly and process actual patient records by using the Uniform Resource Location (URL) and associated parameter passing methods.

The WEB interface to MedLEE has been very well received. It has opened the technology to the research community and has greatly facilitated collaboration within and outside of CPMC.

## Acknowledgments

This publication was supported in part by grant LM05397 from the National Library of Medicine and by the Columbia CAT, a New York State Center for Advanced Technology supported by the New York State Science and Technology Foundation.

## References

- [1] Friedman C, Hripcsak G, DuMouchel W, Johnson SB, and Clayton PD. Natural language processing in an operational clinical information system. *Journal of Natural Language Engineering*, 1995:1(1):83-108.